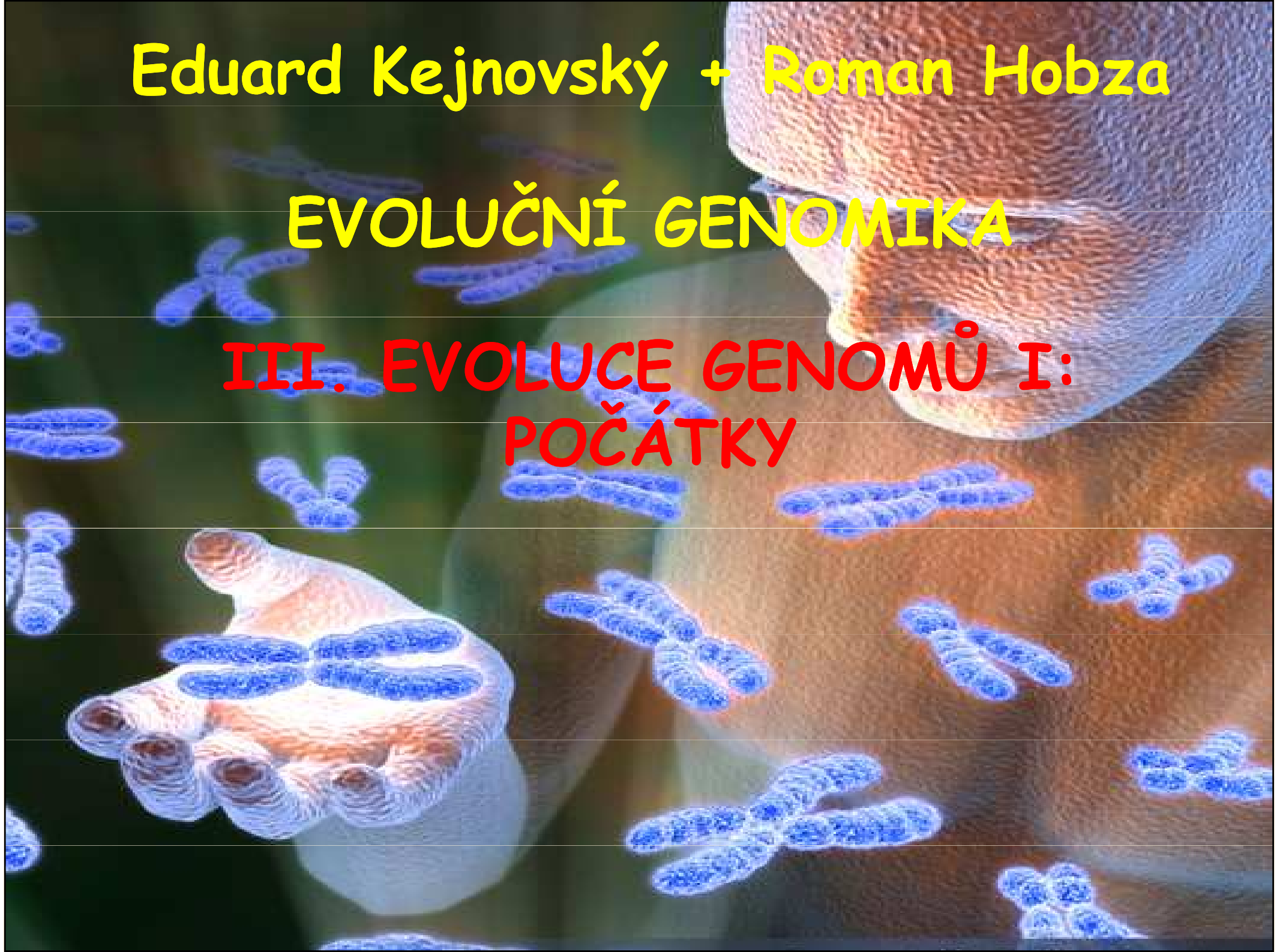


Eduard Kejnovský + Roman Hobza

EVOLUČNÍ GENOMIKA

III. EVOLUCE GENOMŮ I: POČÁTKY



OSNOVA

1. První genomy
2. Původ prokaryot a hypotéza termoredukce
3. Velikost genomu
4. Extrémní genomy

PRVNÍ GENOMY

První protein: RNA-dependentní RNA polymeráza (RNA replikáza)

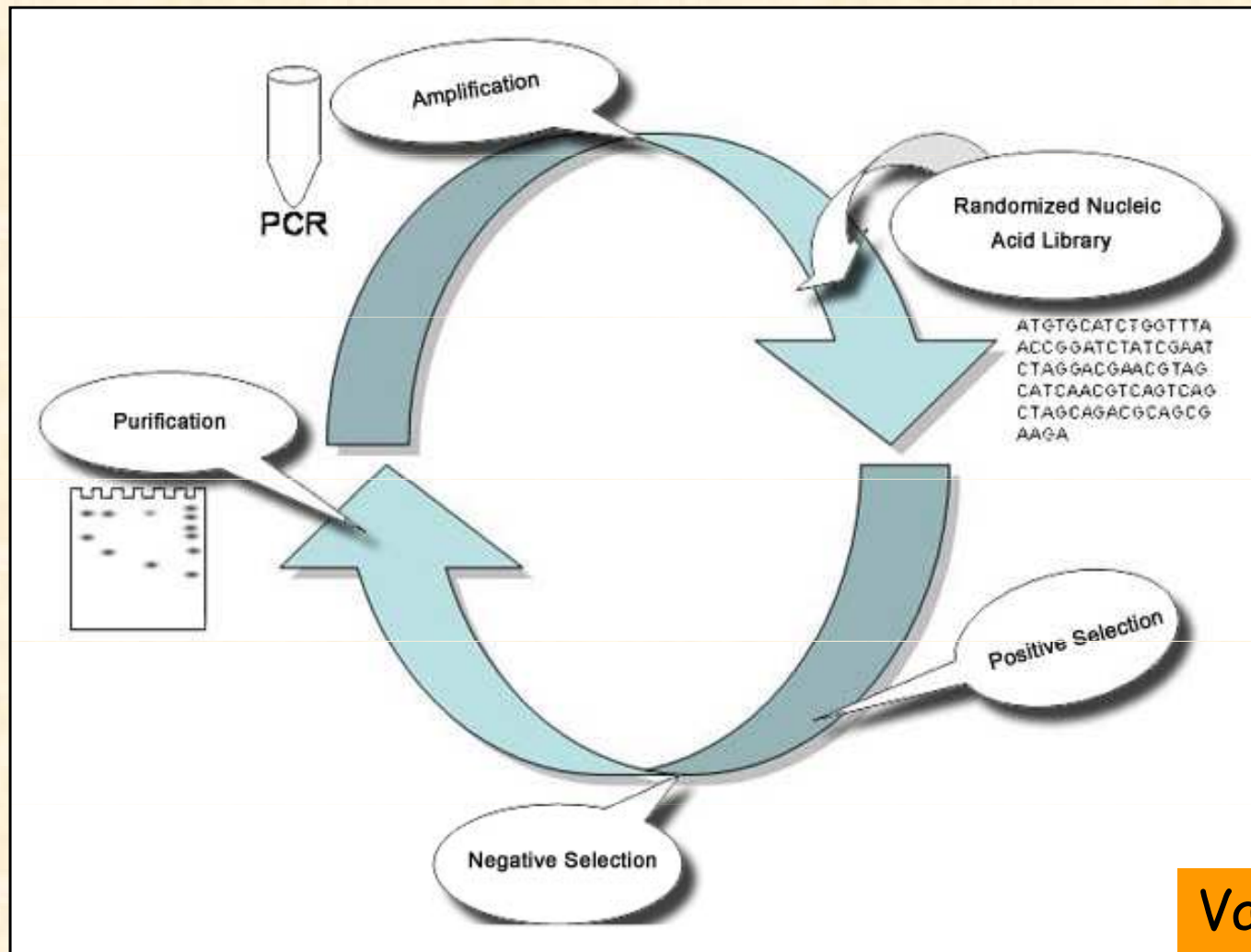
RNA → RNP → protein

Proteiny zvýšily účinnost ribozymů

- první geneticky kódovaný protein vznikl náhodou
- krátký peptid strukturně jednoduchý
- interagoval s RNA replikonem, zvyšoval jeho stabilitu či zlepšoval konformaci
- syntéza potomstva musí být **rychlejší** než degradace rodičů
- dostatečná **přesnost**, ale ne absolutní (možnost evoluce)

RNA polymeráza → Reverzní transkriptáza

Šlechtění RNA ve zkumavce



Fág Q β :

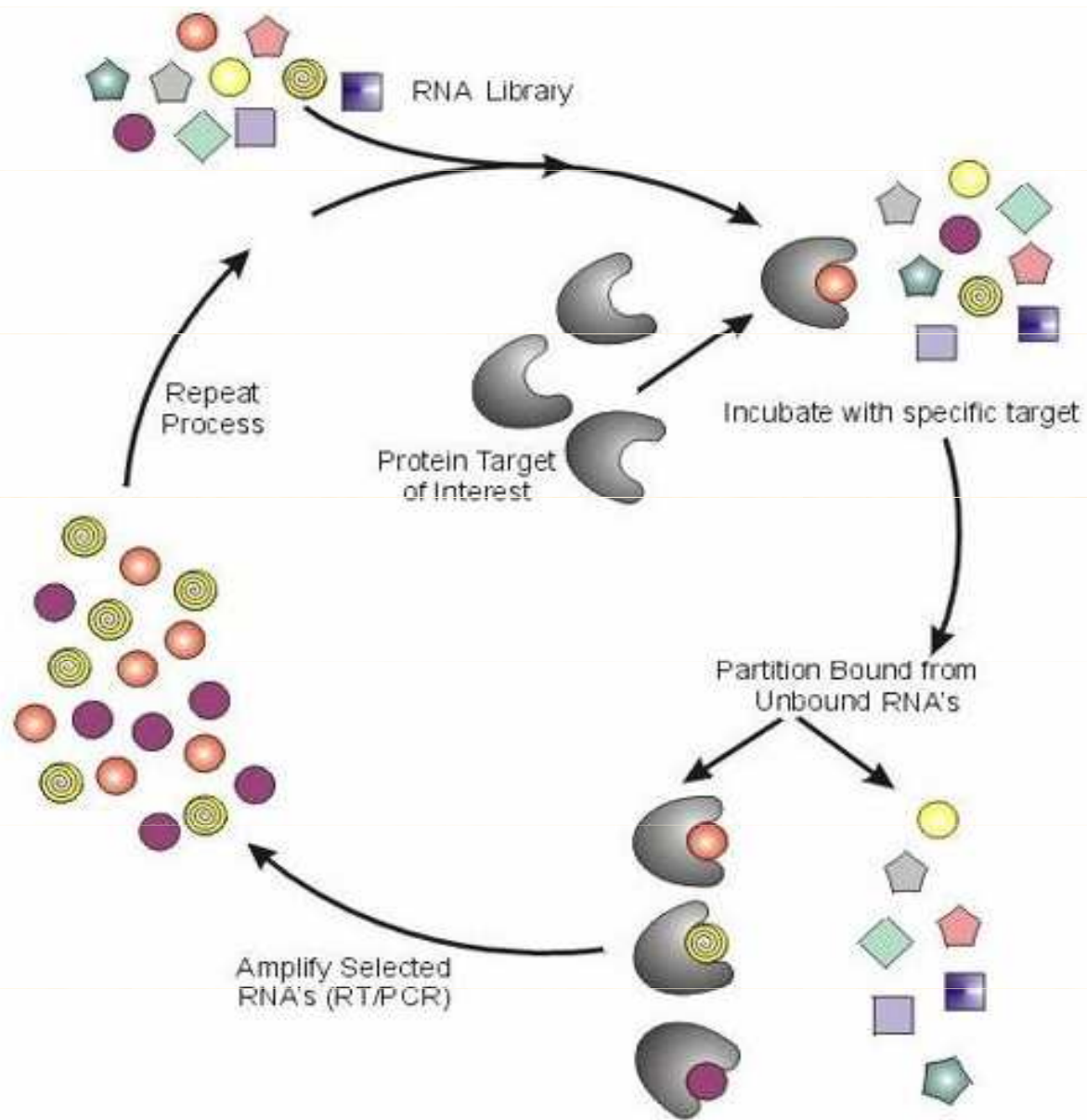
(Spiegelman 1967)

- zkracování času na replikaci --> zkracování RNA genomů (po 74 přenosech eliminace 83% genomů), i jiné selekční tlaky (jedy)

Variabilita/mutace
Selekce
Amplifikace

Tvorba nových proteinů nebo RNA nepřítomných v přírodě s požadovanými vlastnostmi

SELEX: „evoluce *in vitro*“ aneb co by teroristé neměli číst

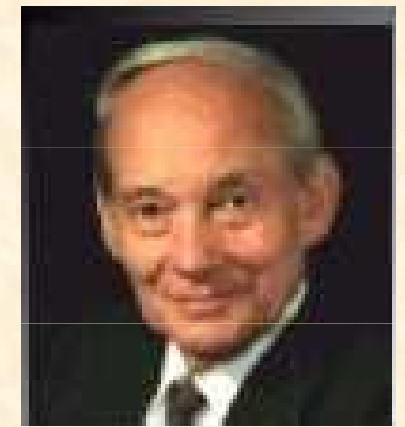


- viry pro genovou terapii
- mutace pláštěvého proteinu
- rezistence k protilátkám
- produkce supervirů

EVOLUCE:

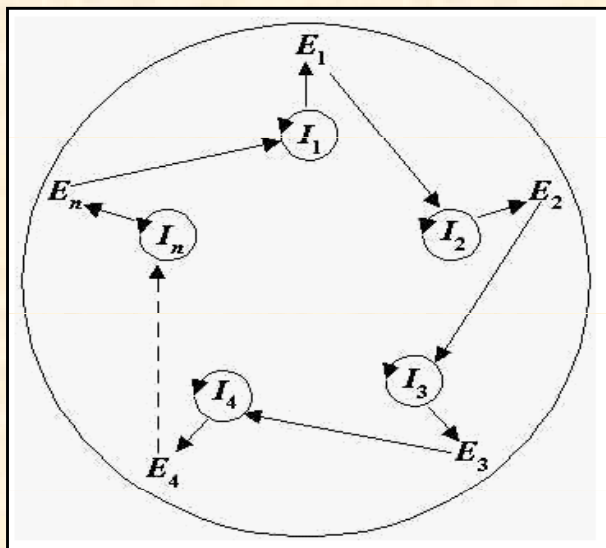
1. dědičnost
2. variabilita
3. selekce

Hypercykly aneb cesta k buňce



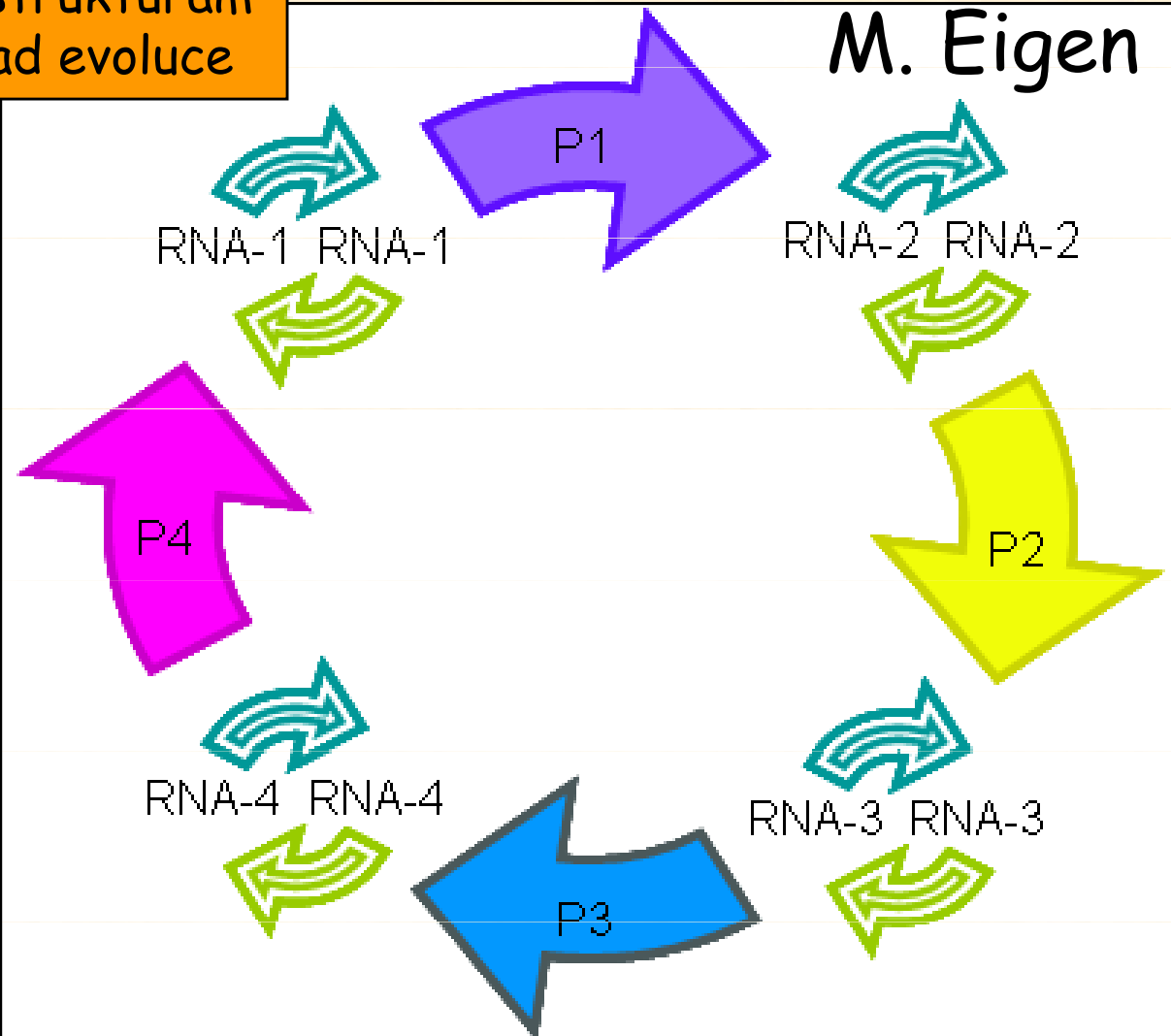
Původně jeden člen - duplikace a mutace - dva členy ...

Cesta od nukleových kyselin ke strukturám buněk = jedna z největších záhad evoluce



V určité fázi vývoje se objevily první **parazité** - zlodějské cykly. Přežily jen hypercykly schopné se bránit parazitům. Za vznikem buněk tedy možná stáli parazité (hybná síla evoluce)

M. Eigen



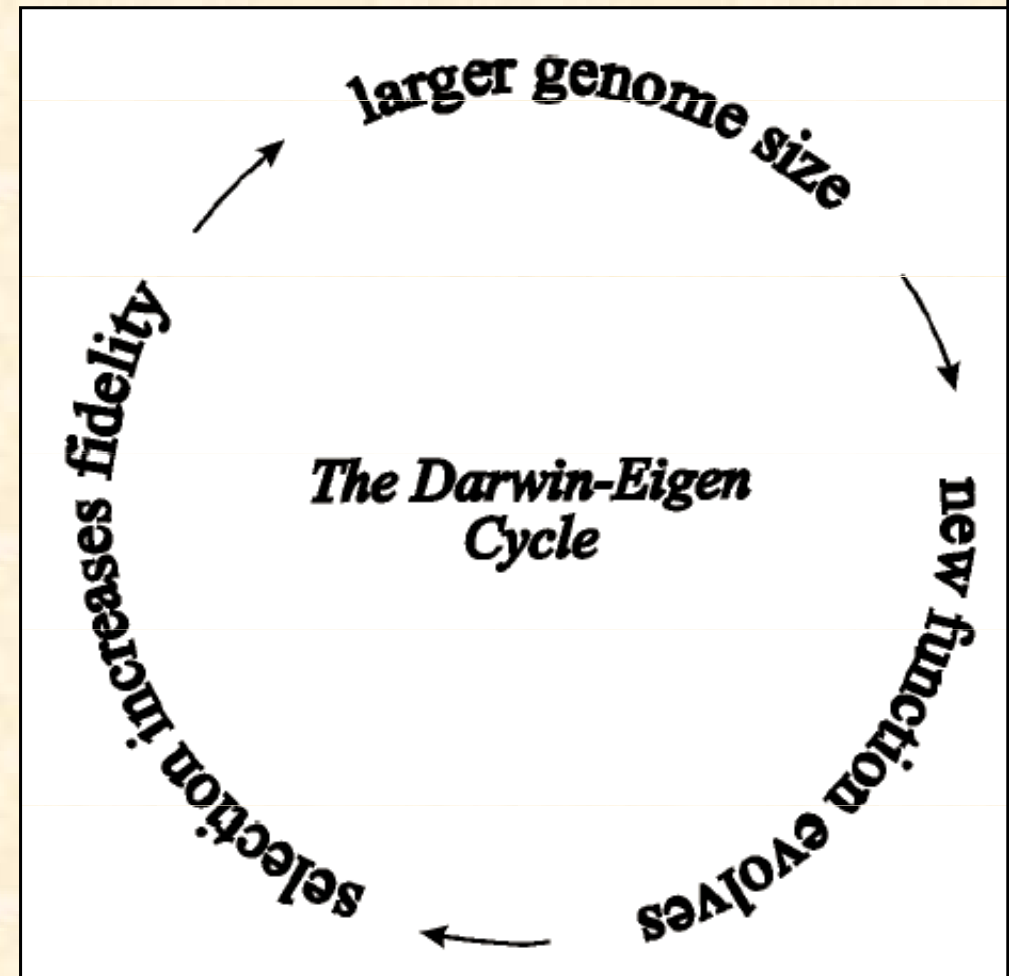
Eigenův limit: replikační přesnost je limitujícím faktorem

- **Definice:**

Čím je vyšší frekvence chyb při replikaci, tím menší genom může projít do další generace

- **Omezení katastrofických dopadů chyb replikace:**

- více kopií (ploidie)
- fragmentace genomu do chromosomů
- rekombinace



První RNA organizmus kódující proteiny: *Riborgis eigensis*

(Jeffares 1998)

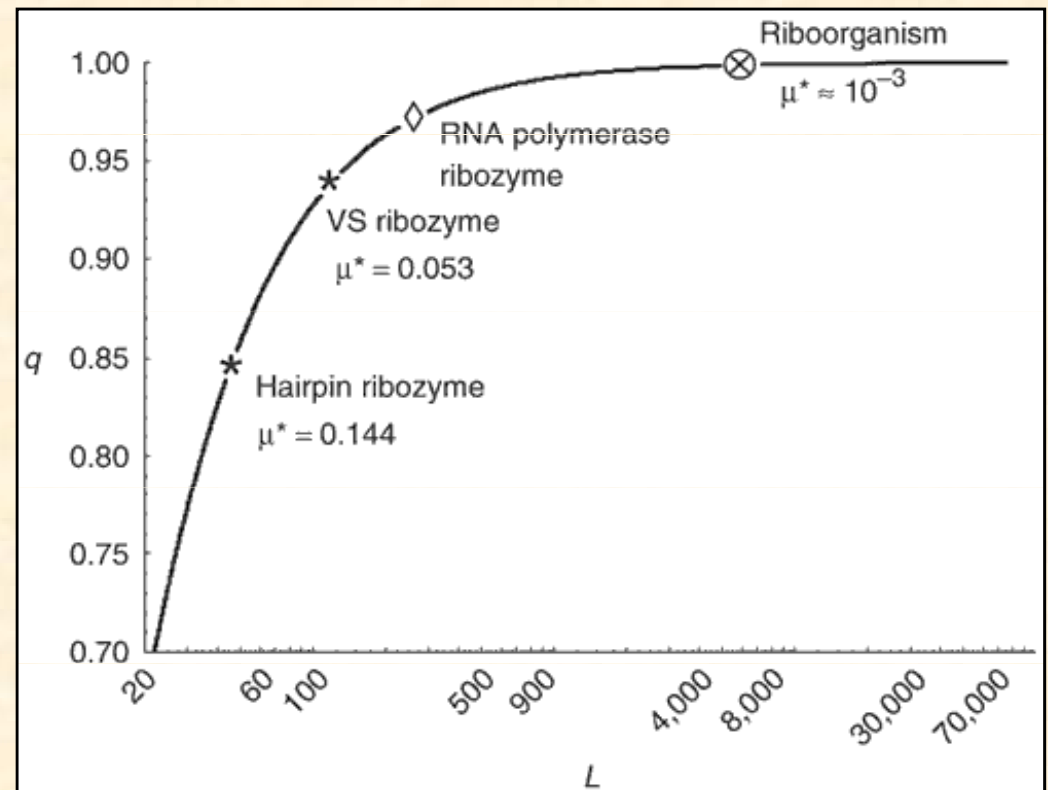
První RNA genomy replikované RNA polymerázami - kódovaly 1 peptidový řetězec

Množství chyb → populace lišících se molekul RNA → koreplikace vzájemně výhodných lineárních molekul kódujících:

- replikázu
- ochranný plášťový protein
- konformační podjednotku.

Vznik fragmentovaných interagujících genomů (podobnost struktury eukaryontního genomu - původní, prokaryota odvozená)

R. eigensis ~ 15kb genom



Dnešní viry: Funkční relikty časných replikonů?

Pohled na viry:

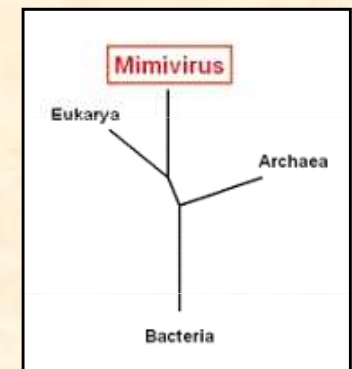
- (a) molekulární paraziti, **odvození** v důsledku způsobu života
- (b) **primitivní**, na hranici života podobně jako časné replikátory
- funkční relikty x funkční modely RNA-proteinových replikonů

RNA viry:

- minimální kódující kapacita (coronaviry 30kb)
- některé viry střídají fáze RNA a DNA - reminiscence RNA→DNA přechodu
- primerem replikace je tRNA

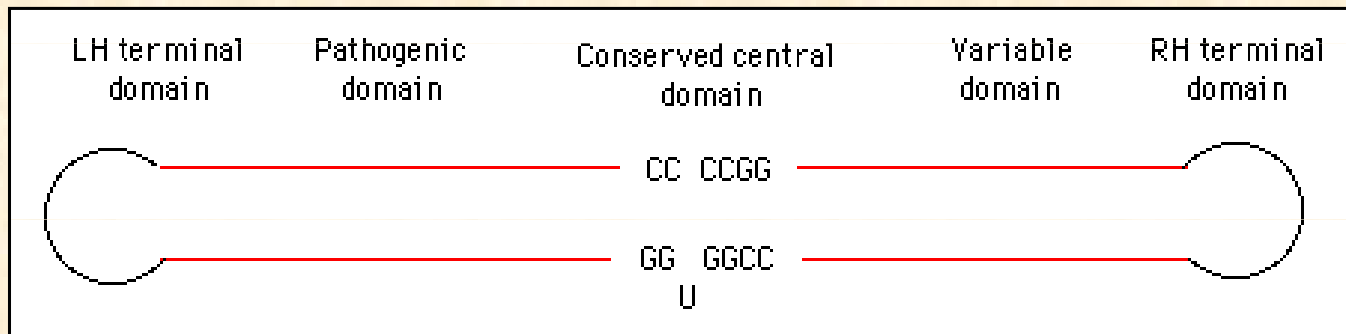
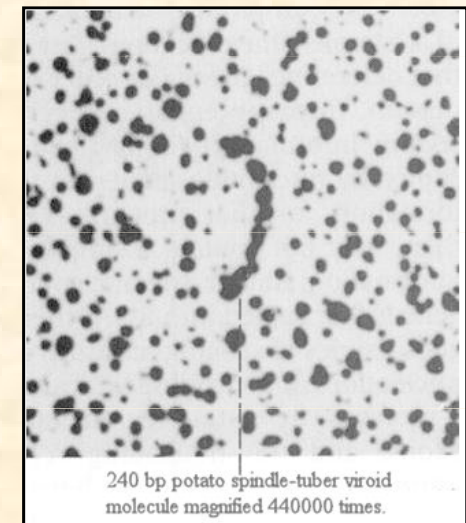
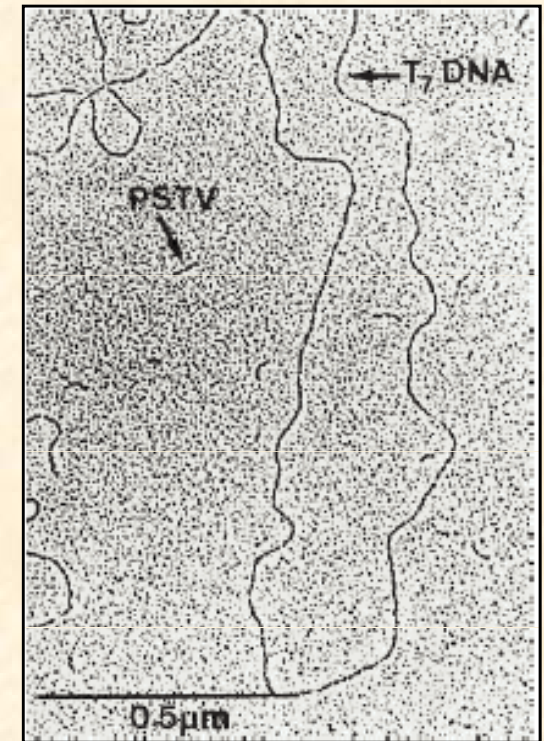
Mimiviry - hranice života:

- velikost genomu srovnatelná s prokaryoty (1.2Mb)
- metabolické geny (911 genů pro proteiny)
- 10% repetitivní DNA
- jen částečná závislost na hostiteli (proteosyntéza)



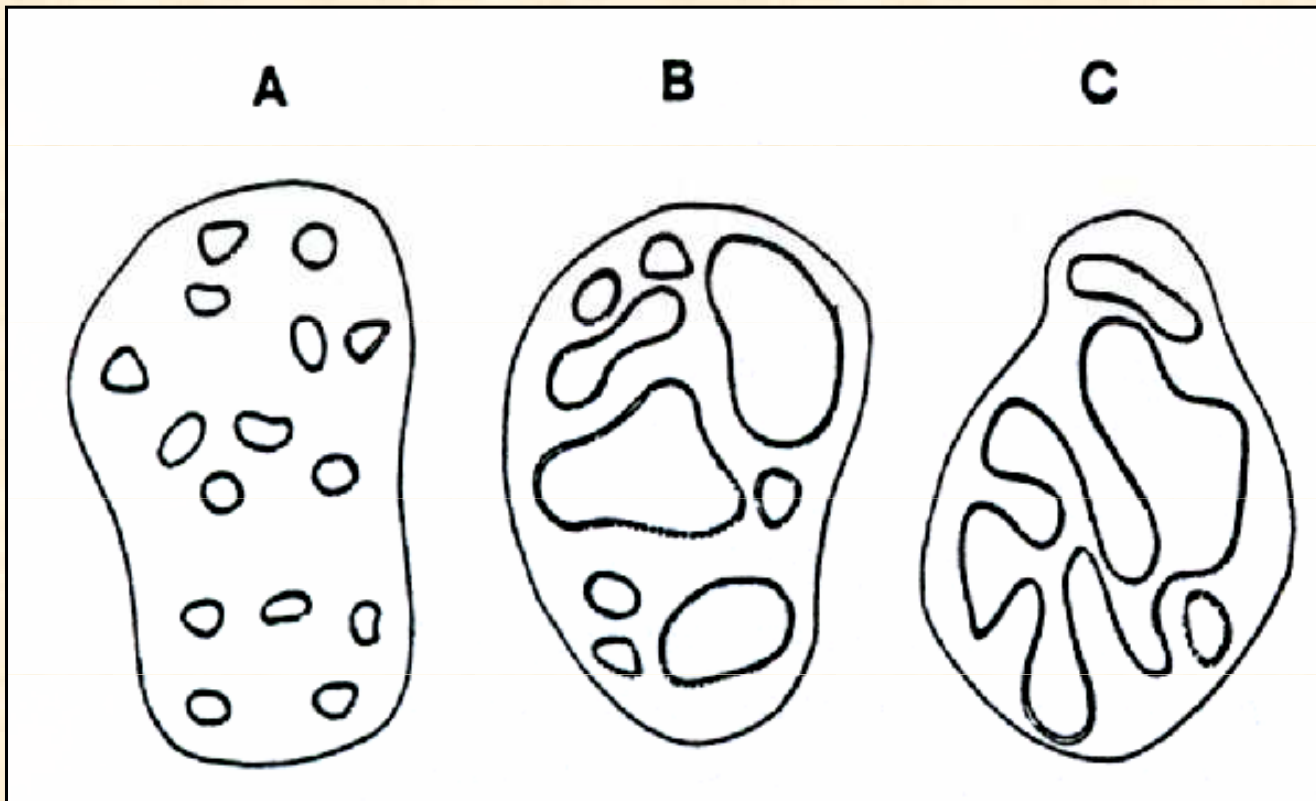
Viroidy: nejpodobnější časným replikonům

- patogeny rostlin
- 200 až 10 000 kopií na buňku
- malé RNA genomy (240-400 b), ssRNA, cirkulární,
- nekódují proteiny - jako replikony éry před proteiny
- replikovány hostitelskými RNA polymerázami
- rolling-circle mechanismus
- multimery štěpené autokatalytickými **ribozymovými** sekvencemi
- intenzivní **vnitřní párování bází** jejich genomické sekvence
- tvorba sekundárních struktur stabilizujících genomy



První DNA genomy: vznik fúzováním malých kružnic DNA

- první genomy: lineární nebo cirkulární?
- malé kružnicové DNA genomy, disperzní genom
- fúzování, geny jako autonomní DNA
- počty kopií statisticky stejné - podobné přenosům plazmidů



Fáze:

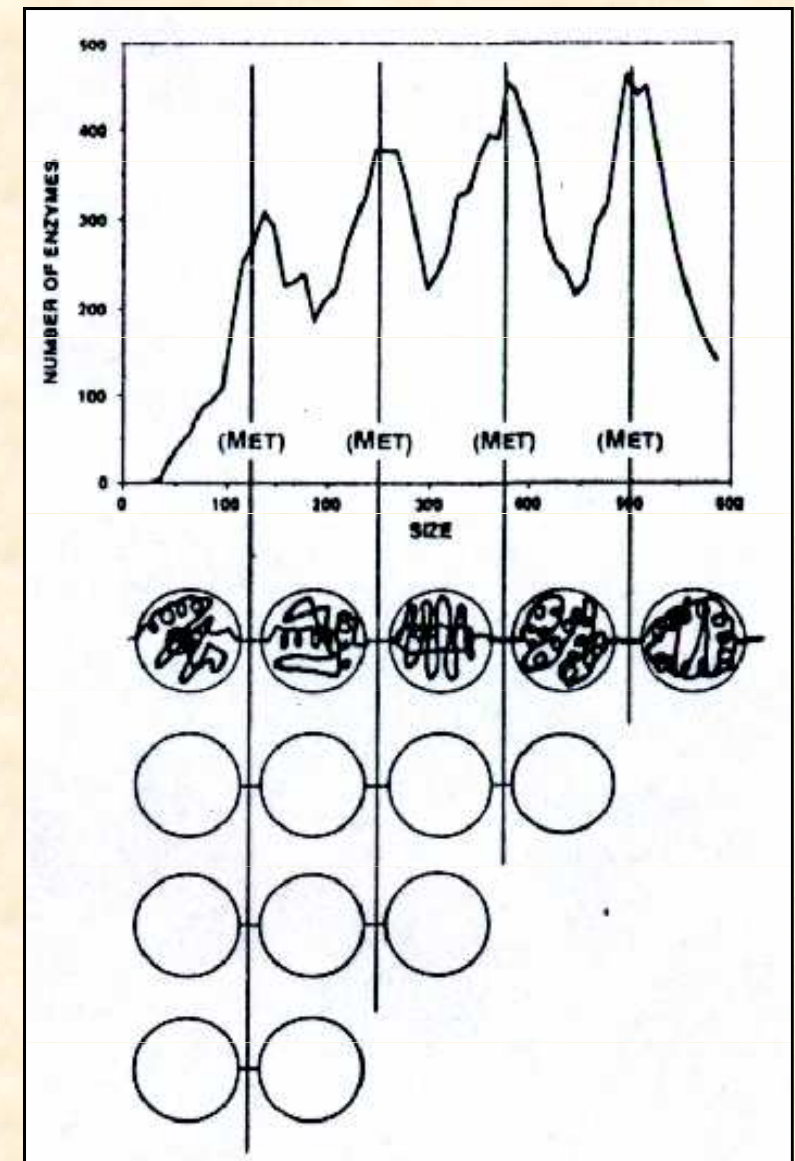
- A. pregenomická
- B. rekombinační
- C. genomická

První genomy - vznik fúzováním malých kružnic DNA

(Trifonov 1995)

Důkazy:

- **periodicita** délek proteinů (123AK u prokaryot, 152 AK u eukaryot)
- nejnazší **cirkularizace** 350-450bp
- periodicita výskytu **Met**
- pozůstatkem **extrachromosomální DNA** - 10% genomu
- **mobilní** elementy, lyzogenní fágy, genomy organel, horizontální přenos
- replikace genů a satelitů prostřednictvím eccDNA



Stromatolity: nejstarší stopy života

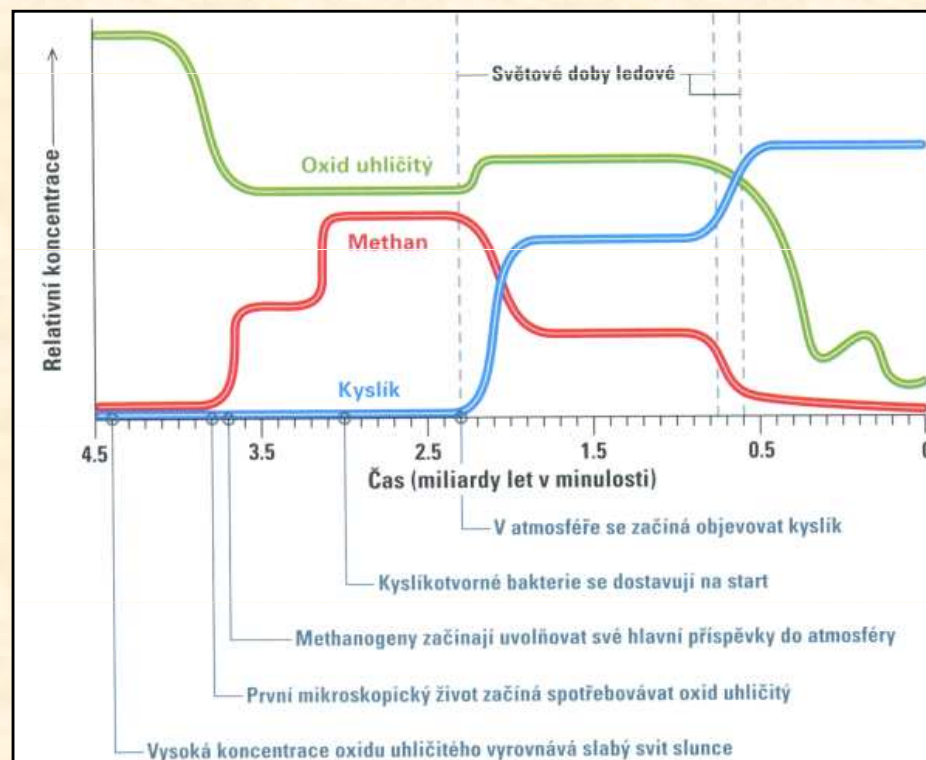
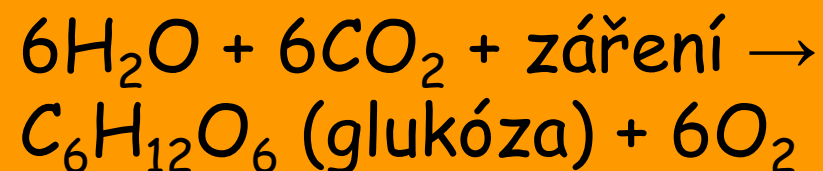
- zkamenělé útvary tvořené sinicemi (?)
- stáří až **3.8 miliardy let**
- anaerobní podmínky (archea) → činností sinic vznikl kyslík
- moderní stromatolity (hypersalinní jezera, Austrálie)



Photo P.-A. Bourque

Organizmy objevily fotosyntézu

- **fotosyntéza** = schopnost tvořit těla z vody a vzduchu za pomoci slunečního záření
- původní organizmy anaerobní, živily se organickými látkami v praprolévce
- **kyslík** jedovatý (i pro dnešní anaerobní organizmy, radikály), nahrazení CO_2 (skleníkový plyn) kyslíkem vedlo k ochlazení planety
- schopnost organismů ovlivňovat klima planety - „**Gaia** živoucí planeta“ (James Lovelock)
- **Symbiogeneze** (Lynn Margulis) - původně parazitismus? (J. Flegr)



POMĚRNÉ KONCENTRACE hlavních plynů ovzduší mohou vysvětlit výskyt celosvětových dob ledových v dávné minulosti Země. Methanogenní mikroorganismy na počátku vzkvétaly, ale když se před 2,3 miliardami let začal z hornin uvolňovat kyslík, zbylo pro tuto stvoření jen několik málo útočišť, kde mohla přežít. Doprovodný pokles koncentrace methanu – silného skleníkového plynu – mohl ochladit celou planetu. Role oxidu uhličitého, nejvýznamnějšího skleníkového plynu v dnešní atmosféře, byla tehdy mnohem méně dramatická.

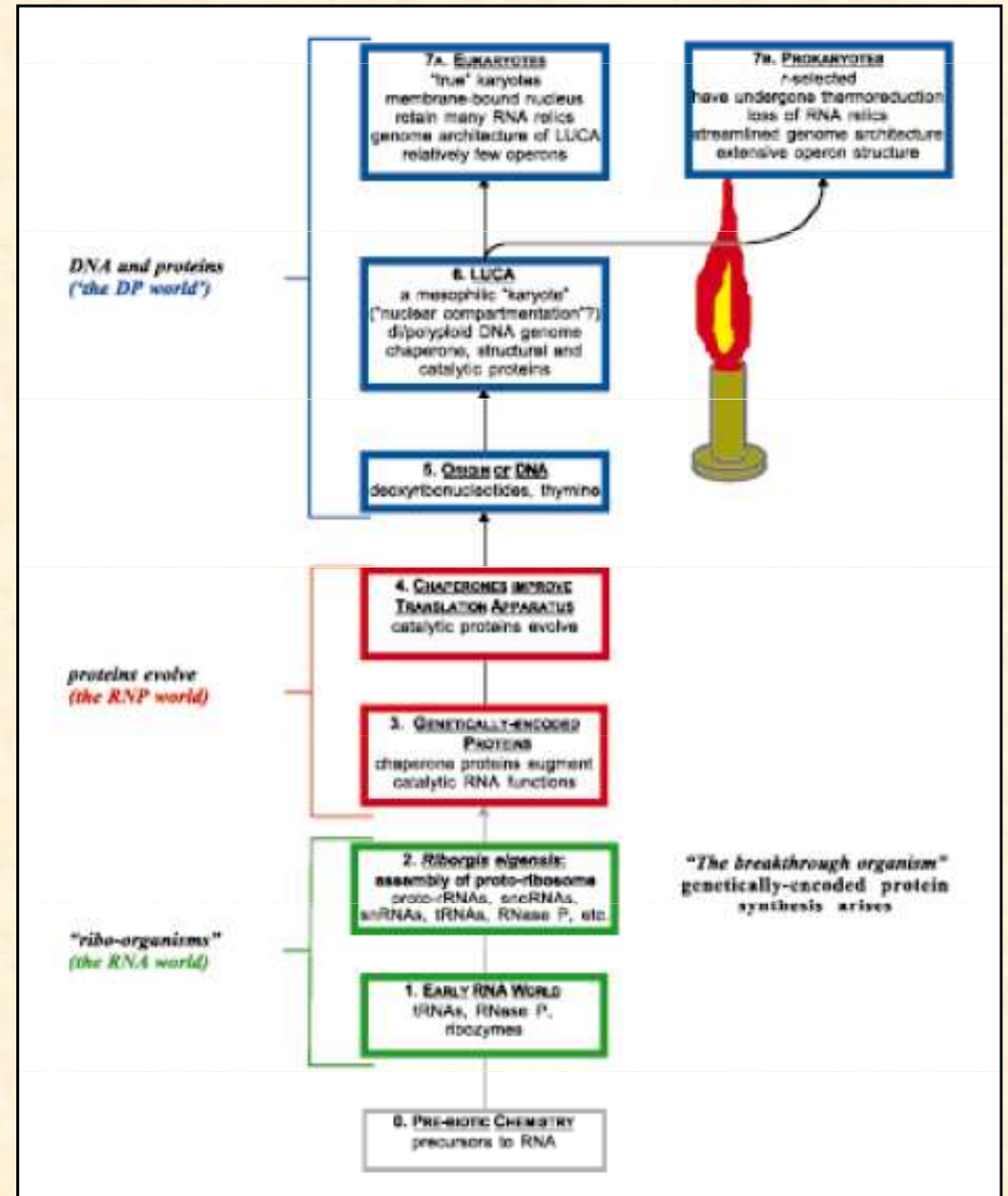
**PŮVOD PROKARYOT
A HYPOTÉZA
TERMOREDUKCE**

Svět RNA podporuje představu starobylosti eukaryot

1. Mnoho reliktnů světa RNA u eukaryot (snRNA, snoRNA, gRNA, telomeráza), jen některé také u prokaryot (RNázaP, tRNA), proč by účinnější proteiny byly nahrazeny molekulami RNA
2. Posttranskripční úpravy mRNA a rRNA jsou rychlé a účinné u prokaryot
3. Neexistuje selekční výhoda pro moderní vznik sestřihu a spliceosomu u eukaryot, složitější struktura, mRNA za 1hod místo za 1 minutu
4. Eukaryotické telomerázy jsou starobylé struktury, homologie s RT

Původ prokaryot a hypotéza „termoredukce“

- v evoluci prokaryot bylo stádium termofilních organizmů
- malá stabilita RNA při teplotách nad 50°C
- časové i prostorové oddělení transkripce a translace nevýhodné - degradace RNA
- odstraněním intronů odpadl náročný sestřih
- malé RNA vymizely nebo nahrazeny stabilnějšími proteiny



Genomy prokaryot jsou mladší a odvozené

Derivation of the Prokaryotes

	<i>Riborgis eigensis</i>	LUCA	Eukary- otes	Prokary- otes
DNA genome	—	+	+	+
Diploid or polyploid	+	+	+	—
Telomerase RNA	+	+	+	—
Linear Genome	+	+	+	—
rRNA processing by snoRNA	+	+	+	—
mRNA processing by snRNA ^a	+	+	+	—
tRNA processing by RNase P	+	+	+	+
Coupled rRNA transcription & processing	—	—	—	+
Coupled transcription & translation	—	—	—	+
Genome-encoded CCA tail ^b	—	—	—	+

Genomy prokaryot jsou mladší a odvozené

PŮVODNÍ GENOM

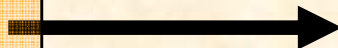
- lineární
- fragmentovaný
- introny obsahující
- RNA molekuly potřebné pro úpravy RNA

EUKARYOTA

ODVOZENÝ GENOM

- cirkulární
- jedna molekula
- operony obsahující
- mnohé RNA nahrazeny proteiny

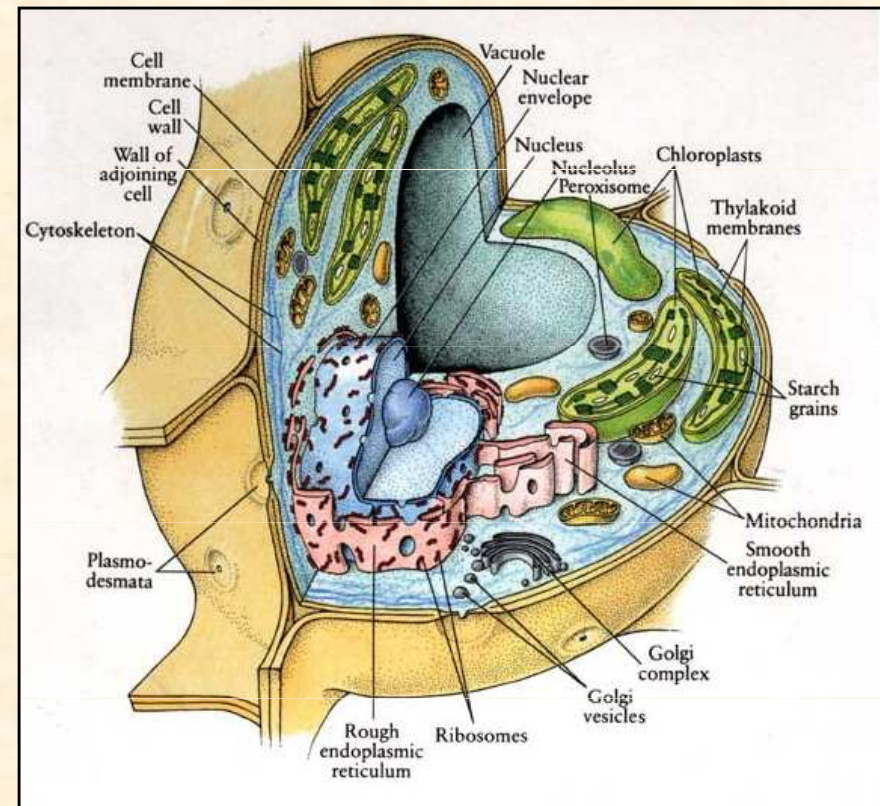
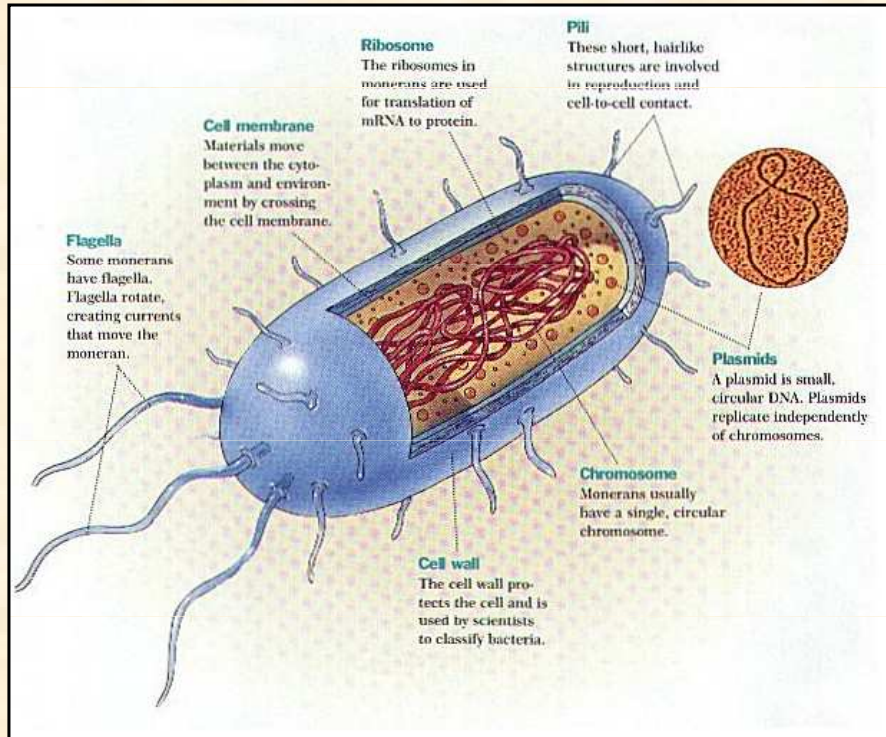
PROKARYOTA



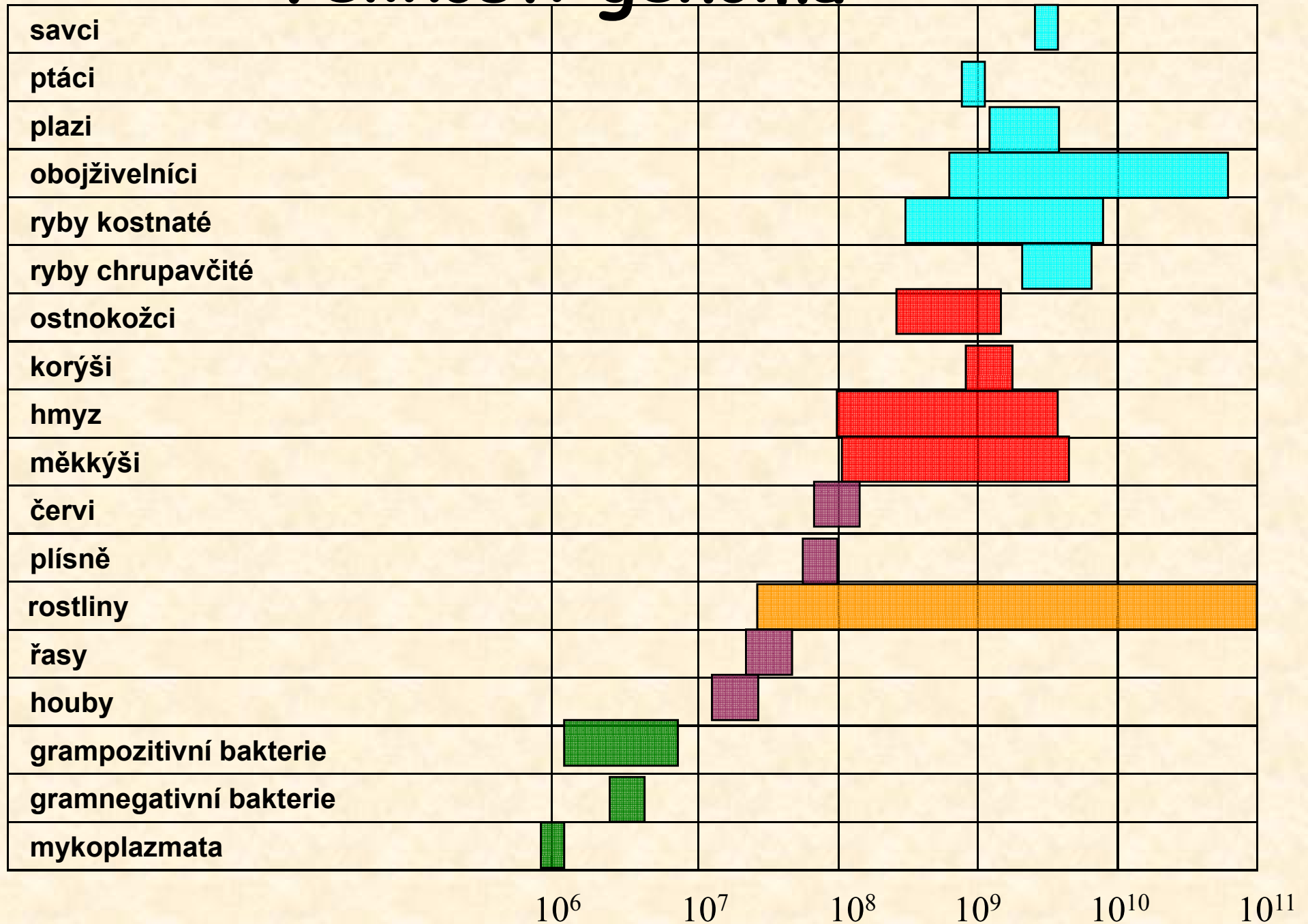
VELIKOST GENOMU

Definice genomu

- celková genetická informace organismu
- u autonomních forem života vždy v DNA
- **prokaryota**: cirkulární chromosom + plazmidy
- **eukaryota**: chromosomy v jádře, mitochondrie a chloroplasty



Velikosti genomů



Paradox hodnoty C

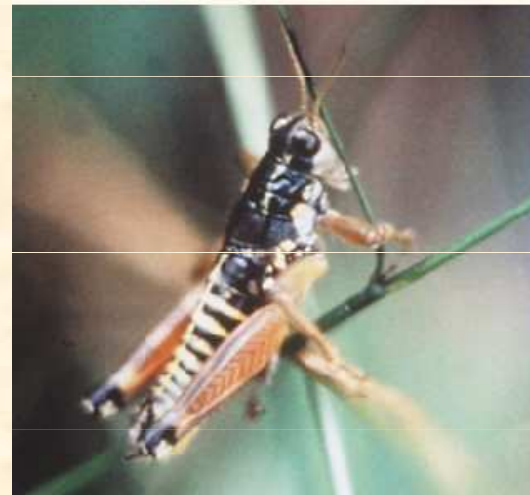
Definice:

- velikost genomu organismu **není v korelaci** s komplexitou organismu (*H. sapiens* 200x menší genom než *Amoeba dubia*)
- obsah kódující DNA podobný, příčinou je **nekódující DNA**, sobecká DNA



180 Mb

Drosophila melanogaster

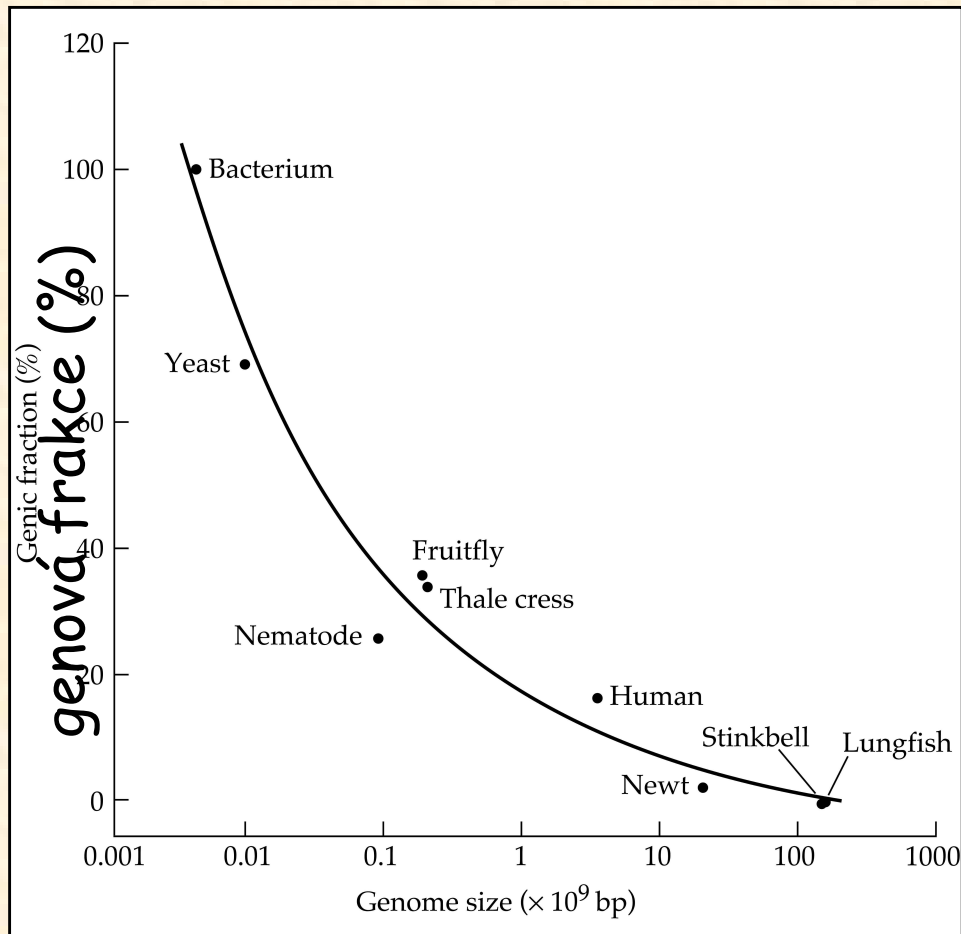


18,000 Mb

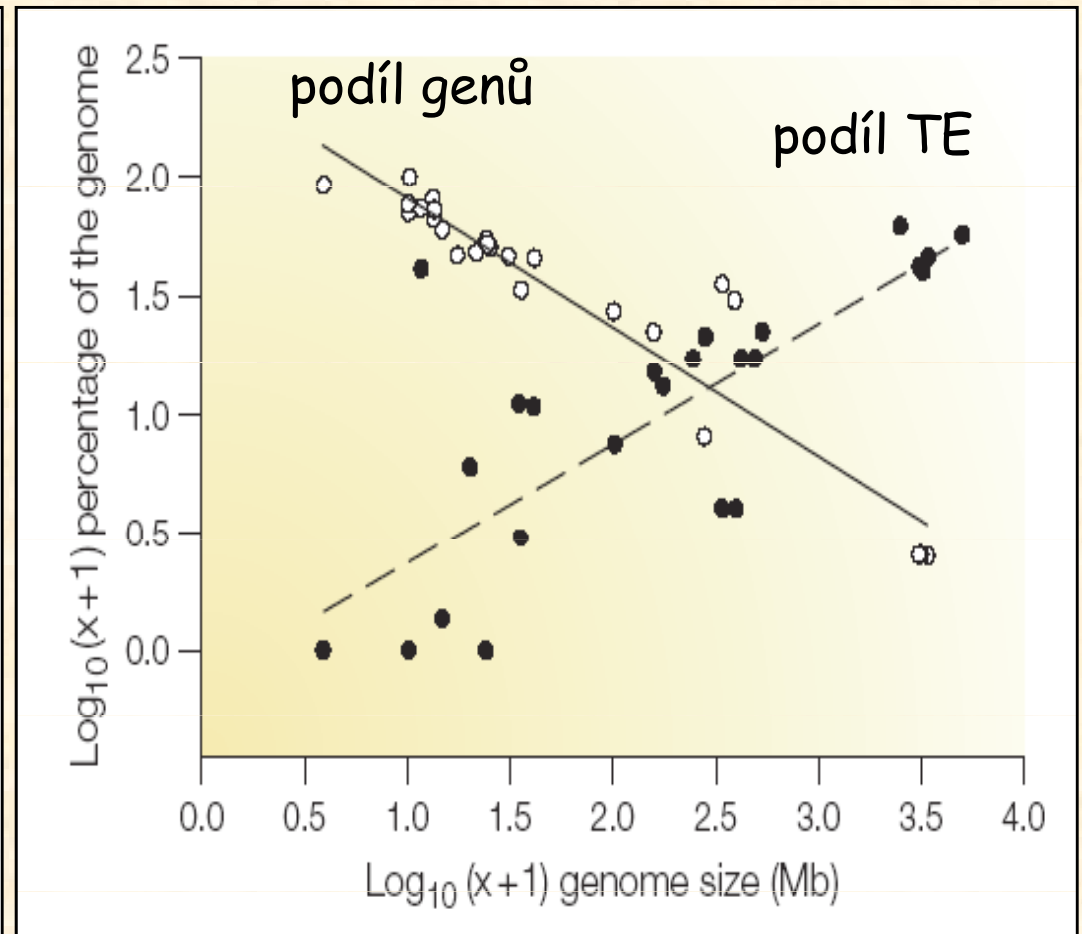
Podisma pedestris

Skutečnost, že genomy eukaryot se liší až 80,000-krát -
snad největší výzva pro genomové holisty

Klíč k řešení paradoxu hodnoty C se nachází v negenových oblastech



velikost genomu



velikost genomu

Hlavní komponenty eukaryotického genomu

Kódující části genů:

- u prokaryot tvoří většinu genomu
- u eukaryot méně, člověk 24 000 genů - 1.5%

Introny:

- původně považovány za příčinu C-paradoxu,
- tvoří většinu genů

Pseudogeny:

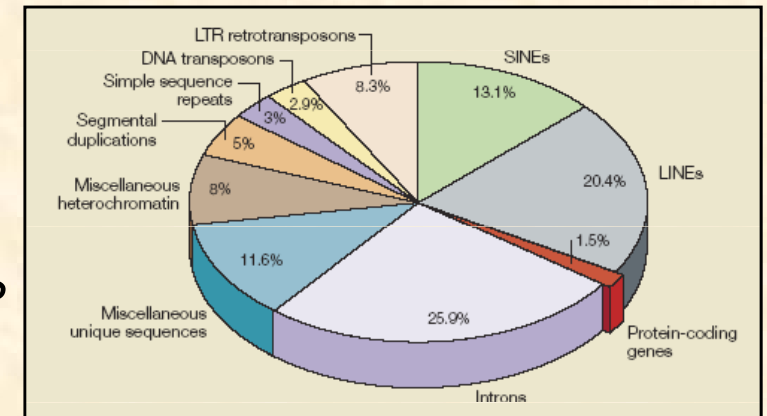
- klasické, retropseudogeny,
- 19000 člověk, 14000 myš, 51 kur, 33 kvasinka, 176 drosophila

Mobilní elementy:

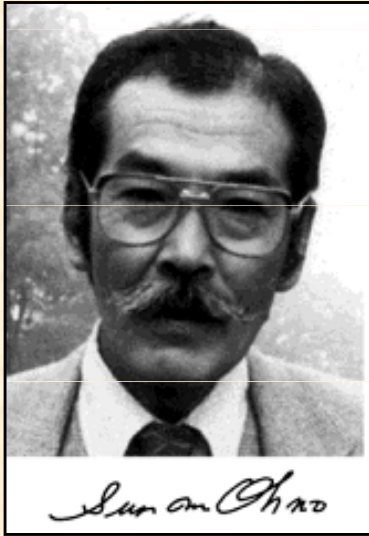
- LTR, nonLTR - SINE, LINE, DNA transposony-MITE

Numt, Nupt:

- inserce promiskuitní DNA



Velikosti genomů a zbytečná DNA



„The majority of the genome consisted of now-extinct genes“

- Susumu Ohno

“Triumphs as well as failures of nature's past experiments appear to be contained in our genome“

- Susumu Ohno

„The evolution of the large-scale features of the genome is one of the most difficult, perhaps the most difficult, question in evolutionary biology“

- Maynard Smith

Velikosti genomů prokaryot

- **Bakteriální genom: 6×10^5 - 10^7 (=0.6-10Mb)**
nejmenší známý: *Mycoplasma genitalium* (480 genů kódujících proteiny, 3 rRNA geny, 37 tRNA genů)
- **Procesy ovlivňující velikost bakteriálního genomu:**
Genová duplikace, malé delece a inserce, transpozice, horizontální přenos, ztráta genů v parazitických liniích, atd.

Velikost genomů prokaryot je zhruba úměrná počtu genů

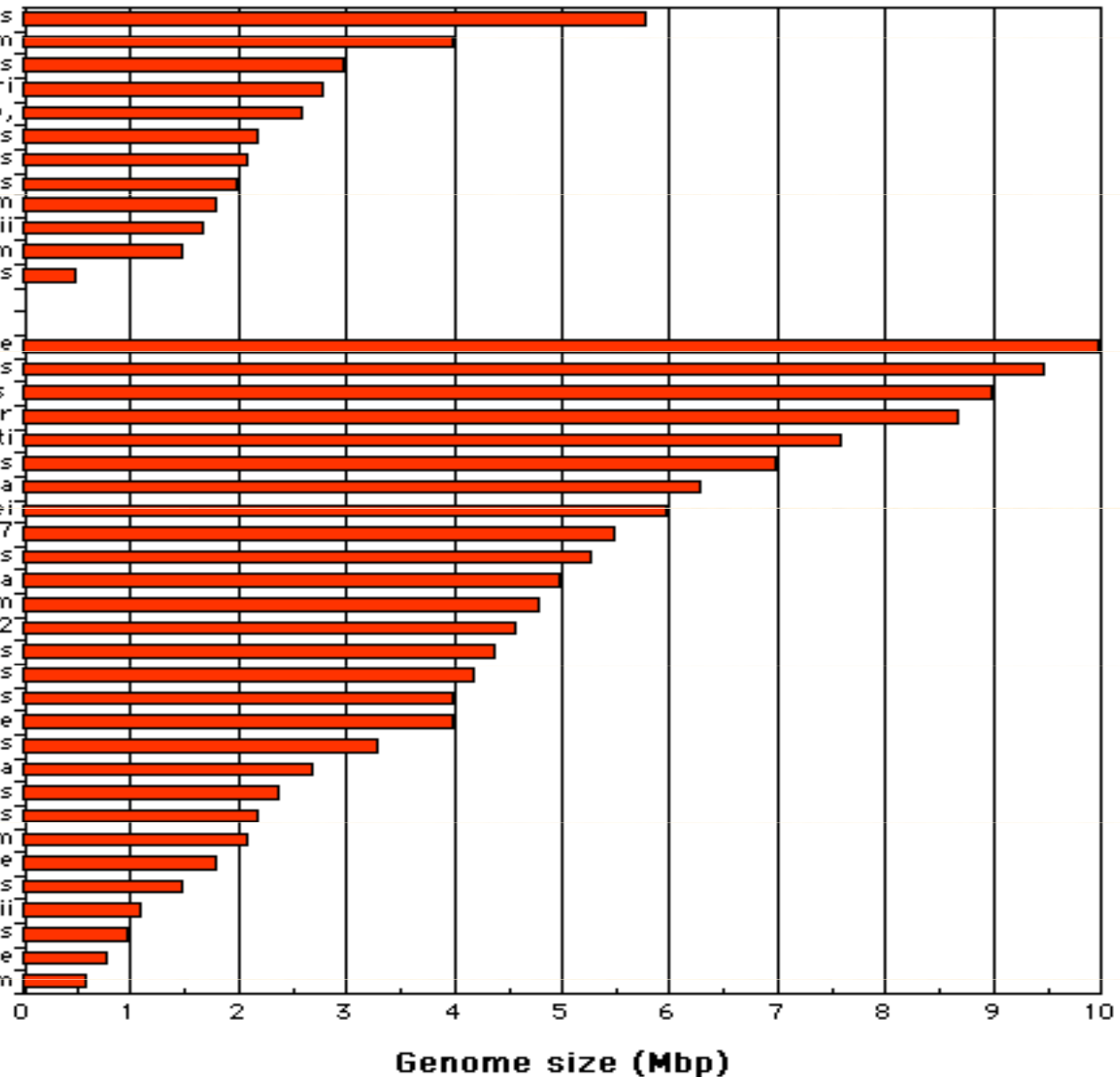
Velikosti genomů prokaryot

Archaea:

Methanosarcina acetivorans
Halobacterium salinarium
Sulfolobus solfataricus
Methanosarcina barkeri
Halobacterium sp,
Archaeoglobus fulgidus
Pyrococcus furiosus
Ferroplasma acidarmanus
Methanobacterium thermoautotrophicum
Methanococcus jannaschii
Thermoplasma acidophilum
Nanoarchaeum equitans

Bacteria:

Nostoc punctiforme
Myxococcus xanthus
Gemmata obscuriglobus
Streptomyces coelicolor
Mesorhizobium loti
Mycobacterium smegmatis
Pseudomonas aeruginosa
Burkholderia pseudomallei
Escherichia coli O157:H7
Agrobacterium tumefaciens
Pseudomonas putida
Salmonella typhimurium
→ Escherichia coli K-12
Mycobacterium tuberculosis
Bacillus subtilis
Caulobacter crescentus
Vibrio cholerae
→ Deinococcus radiodurans
Xylella fastidiosa
Lactococcus lactis
Neisseria meningitidis
Chlorobium tepidum
→ Haemophilus influenzae
Aquifex aeolicus
Rickettsia prowazekii
Geobacter sulfurreducens
Mycoplasma pneumoniae
→ Mycoplasma genitalium




EXTRÉMNÍ GENOMY

Minimální velikost genomu

„One day a scientist will drop gene number 297 into a test tube, then number 298, then 299... and presto: what was not alive a moment ago will be alive now. The creature will be as simple as life can be. But it will still be life. And humans will have made it, in an ordinary glass tube, from off-the-shelf chemicals...”

- Hayden, 1999:



TIGR
THE INSTITUTE FOR GENOMIC RESEARCH

Minimal Genome Project

Press Release

Co je život?
Strategie: 1. odstranění či inaktivace genů
2. syntéza minimálního genomu

TIGR Databases
What's New
About TIGR
TIGR Faculty

2165-2169 (1999)

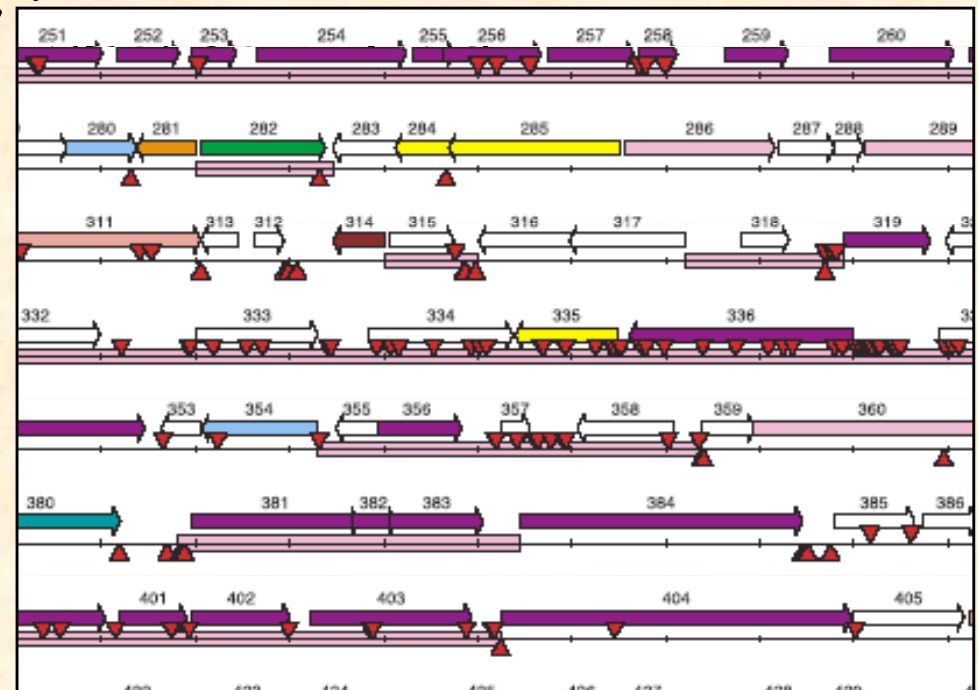
"Minimální" genom - Mycoplasma

580 kb genom/ 480 genů pro proteiny/ 37 genů pro tRNA
2209 inzercí transpozonů/ ve 140 genech
1354 míst, kde inzerce nebyla letální
265-350 genů nepostradatelných (glykolýza)
180-215 genů postradatelných
100 genů má neznámou funkci!!!
různý vliv inzercí podle polohy

Global Transposon Mutagenesis and a Minimal Mycoplasma Genome

Clyde A. Hutchison III,^{1,2*} Scott N. Peterson,^{1†} Steven R. Gill,¹
Robin T. Cline,¹ Owen White,¹ Claire M. Fraser,¹
Hamilton O. Smith,^{1‡} J. Craig Venter^{1‡§}

Mycoplasma genitalium with 517 genes has the smallest gene complement of any independently replicating cell so far identified. Global transposon mutagenesis was used to identify nonessential genes in an effort to learn whether the naturally occurring gene complement is a true minimal genome under laboratory growth conditions. The positions of 2209 transposon insertions in the completely sequenced genomes of *M. genitalium* and its close relative *M. pneumoniae* were determined by sequencing across the junction of the transposon and the genomic DNA. These junctions defined 1354 distinct sites of insertion that were not lethal. The analysis suggests that 265 to 350 of the 480 protein-coding genes of *M. genitalium* are essential under laboratory growth conditions, including about 100 genes of unknown function.



“Minimální” genom: *Bacillus subtilis* život s 271 geny

4100 genů

192 genů zcela **nepostradatelných**
na bohatém prostředí: zpracování
informace, buněčná stěna,
energie

79 genů **podstatných**

Jen 4% genů má **neznámou** funkci

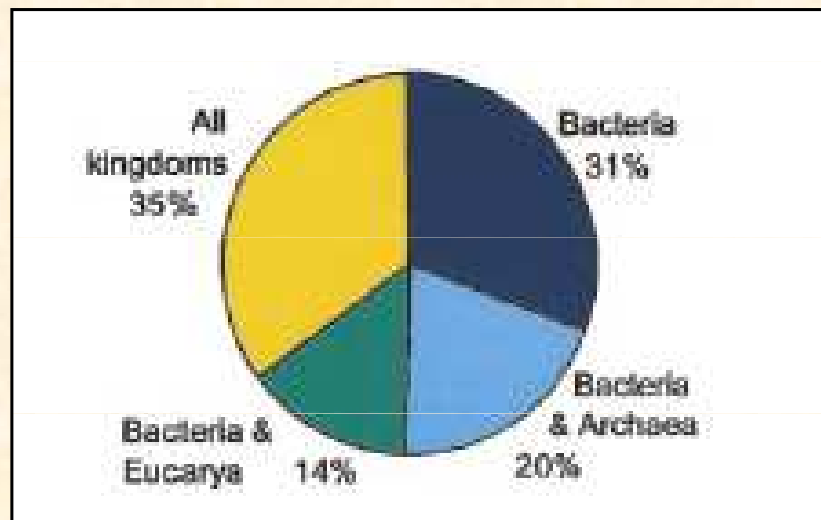


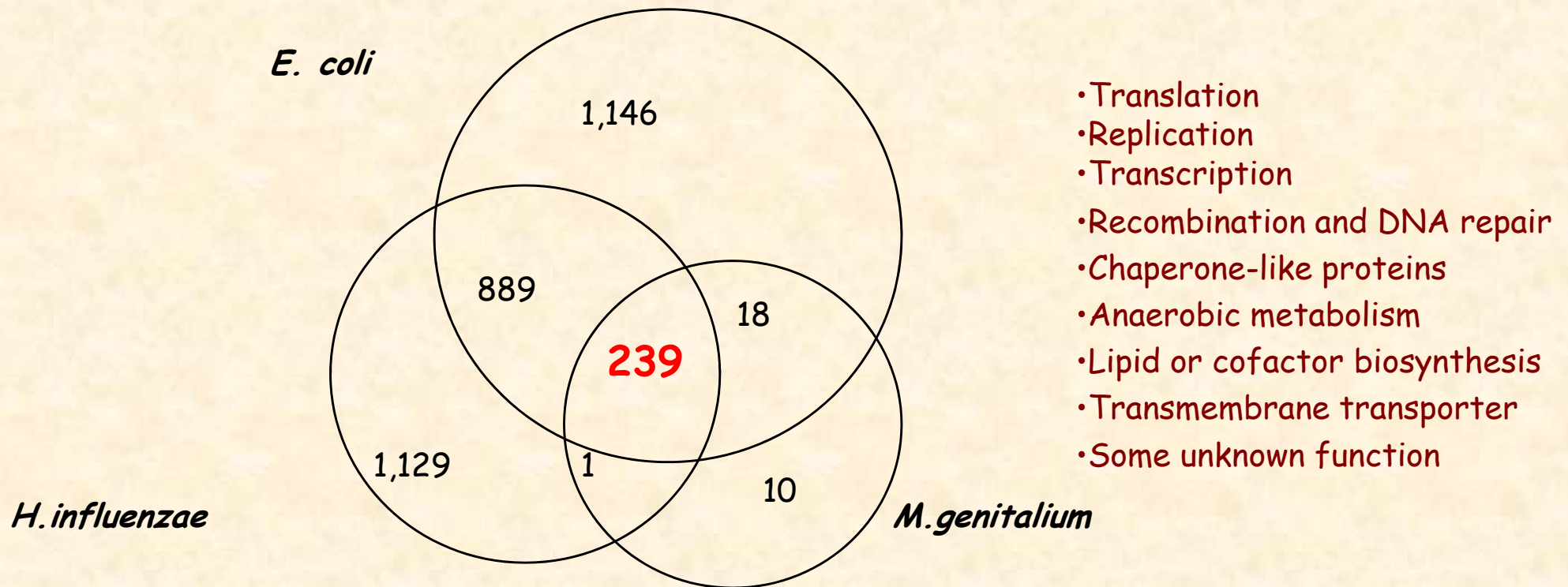
Table 2. *B. subtilis* essential genes

DNA metabolism	27
Basic replication machinery	16
Packaging and segregation	9
Methylation	2
RNA metabolism	14
Basic transcription machinery	4
RNA modification	6
Regulation	4
Protein synthesis	95
Ribosomal proteins	52
Aminoacyl-tRNA synthetases	24
Translation factors	10
Protein folding and modification	3
Protein translocation	6
Cell envelope	44
Membrane lipids	16
Cell wall	28
Cell shape and division	10
Glycolysis	8
Respiratory pathways	22
Isoprenoids	8
Menaquinone	8
Cytochrome biogenesis	3
Thioredoxin	3
Nucleotides	10
Cofactors	15
CoA	1
Folate	3
NAD	4
S-Adenosylmethionine	1
Iron-sulfur cluster	6
Other	15
Unknown	11
Total	271

Minimální velikost genomu: (a) Analytický přístup

(Mushegian a Koonin, 1996):

Srovnání kompletních bakteriálních genomů:



[Překrývající se ortologní geny (239)] + [přesun neortologních genů] - [geny specifické pro parazitické bakterie nebo funkční redundance] = 256 genes

Minimální velikost genomu:

(b) Experimentální přístup

(Itaya, 1995):

Knock-out 79 náhodně vybraných genů *Bacillus subtilis*:

- pouze 6 letálních, 73 je postradatelných
→ 7.5% (6/79) genomu je nepostradatelných
- *B.subtilis* genome: $4.2 \times 10^6 \text{bp} \times 7.5\% = 3.2 \times 10^5 \text{bp}$
Průměrná velikost genu je 1.25kb, takže minimální velikost genomu ≈ 254 genů

Největší prokaryotické genomy

Pseudomonas aeruginosa (bakterie):

- 5500 genů
- přes 6 Mb
- přirozeně rezistentní k antibiotikům (ochranný obal)
- R-faktor, žije ve společenství jiných bakterií, konjugace
- lidský patogen (kožní n., močové, dýchací a trávicí cesty)



Nostoc punctiforme (sinice):

- 7432 ORF
- 8.9 Mb
- **repetice**, transpozony
- fotoautotrofní, také fakultativně heterotrofní
- možnost symbiózy s rostlinami i houbami

Repetice <i>N.p.</i>	Sites in genome
AATGAC _n (STRR2)	
AATGACA	69
AATGACT	63
AATGACC	39
AATTCCC (STRR4)	
AATTCCC	41
AATGCCC	37
AATTACG (STRR5)	45
AATCCCC (STRR1)	
ATCCCC	39
AATCCCC	19
AGTCCCC	15
AGCAGGGG (STRR6)	29
AAAATTC (STRR7)	13

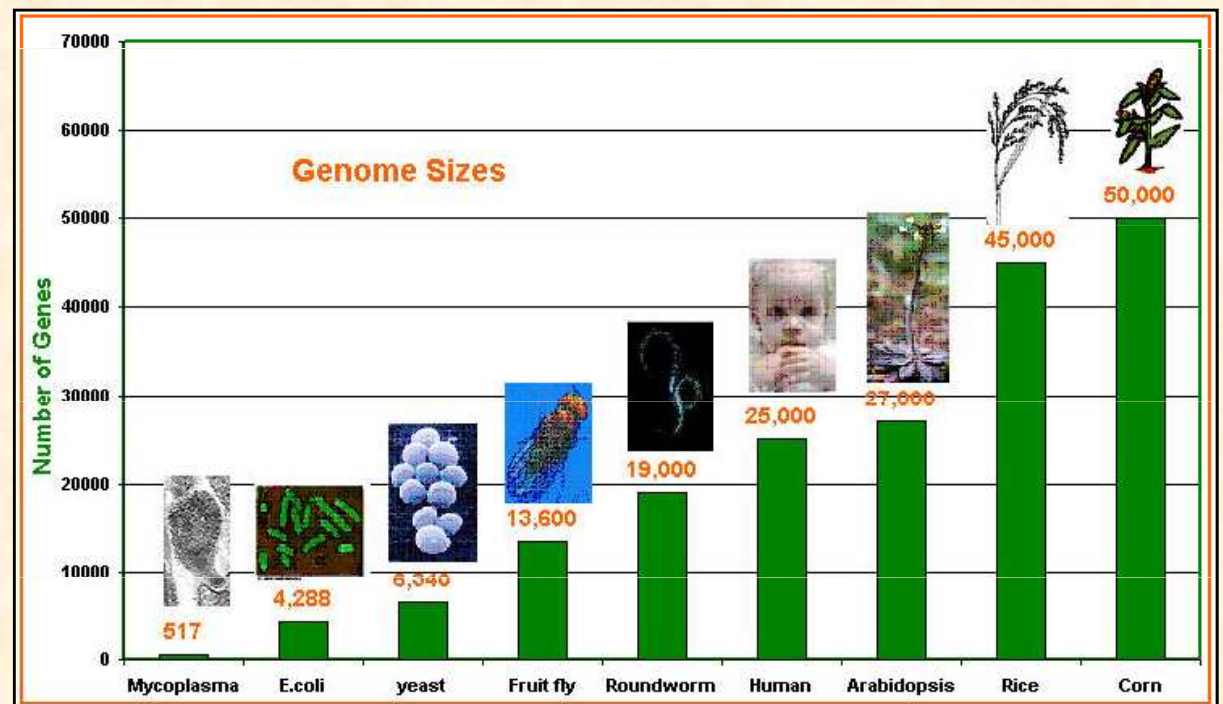
Velikosti genomů eukaryot

- Eukaryotický genom: $8.8 \times 10^6 \sim 6.9 \times 10^{11}$
(8.8Mb - 670Gb)

Nejmenší známý: *Saccharomyces cerevisiae*

Největší známý: *Amoeba dubia*

Velikosti eukaryotických genomů NEjsou úměrné počtu genů nebo komplexitě organismu



Organism	Genome size (base pairs)	Note
Virus, Bacteriophage MS2	3,569	First sequenced RNA-genome ^[2]
Virus, SV40	5,224	^[3]
Virus, Phage Φ -X174;	5,386	First sequenced DNA-genome ^[4]
Virus, Phage λ	50,000	
Bacterium, <i>Haemophilus influenzae</i>	1,830,000	First genome of living organism, July 1995 ^[5]
Bacterium, <i>Carsonella ruddii</i>	160,000	Smallest non-viral genome. ^[6]
Bacterium, <i>Buchnera aphidicola</i>	600,000	
Bacterium, <i>Wigglesworthia glossinidia</i>	700,000	
Bacterium, <i>Escherichia coli</i>	4,000,000	^[7]
Amoeba, <i>Amoeba dubia</i>	670,000,000,000	Largest known genome. ^[8]
Plant, <i>Arabidopsis thaliana</i>	157,000,000	First plant genome sequenced, Dec 2000. ^[9]
Plant, <i>Genlisea margaretae</i>	63,400,000	Smallest recorded flowering plant genome, 2006. ^[9]
Plant, <i>Fritillaria assyrica</i>	130,000,000,000	
Plant, <i>Populus trichocarpa</i>	480,000,000	First tree genome, Sept 2006
Yeast, <i>Saccharomyces cerevisiae</i>	20,000,000	^[10]
Fungus, <i>Aspergillus nidulans</i>	30,000,000	
Nematode, <i>Caenorhabditis elegans</i>	98,000,000	First multicellular animal genome, December 1998 ^[11]
Insect, <i>Drosophila melanogaster</i> aka Fruit Fly	130,000,000	^[12]
Insect, <i>Bombyx mori</i> aka Silk Moth	530,000,000	
Insect, <i>Apis mellifera</i> aka Honey Bee	1,770,000,000	
Fish, <i>Tetraodon nigrovindis</i> , type of Puffer fish	385,000,000	Smallest vertebrate genome known
Mammal, <i>Homo sapiens</i>	3,200,000,000	
Fish, <i>Protopterus aethiopicus</i> aka Marbled lungfish	130,000,000,000	Largest vertebrate genome known